Appl. No. 09/845,561
Amdt. dated July 20, 2004
Reply to Office Action of April 20, 2004

Remarks

The present amendment responds to the Official Action dated April 20, 2004. The

Official Action allowed claims 20 and 22. The Official Action rejected claims 1-9, 23, 27 and 33

under 35 U.S.C. 102(e) based on Ladd U.S. Patent No. 6,493,673 ("Ladd"). The Official Action

rejected claims 10-13 under 35 U.S.C. 103(a) as unpatentable over Ladd. The Official Action

rejected claim 21 under 35 U.S.C. 103(a) as unpatentable over Tanenblatt U.S. Patent No.

6,006,187 ("Tanenblatt") in view of Baba U.S. Patent No. 6,397,183 ("Baba"). The Official

Action objected to claims 14-19, 24, 28, 31, 32 and 34 as dependent on rejected base claims, but

stated that these claims would be allowable if rewritten in independent form including all of the

limitations of the base claim and any intervening claims. These grounds of rejection are

addressed below following a brief discussion of the present invention to provide context. Claims

25, 26, 29 and 30 have been canceled. Claims 1, 21 and 23 have been amended to be more clear

and distinct. Claims 1-24, 27, 28 and 31-34 are presently pending.

The Present Invention

A system according to an aspect of the present invention includes the generation and

processing of a set of tags which can be used to model phenomena. A set of tags can be

developed to represent characteristics of specific phenomena, and the tags may be placed in a set

of instructions for modeling phenomena in order to specify desired characteristics of the

phenomena. The tags may suitably be generated by analyzing instances of actual phenomena in

order to identify characteristics exhibited by the phenomena, and creating a set of tags defining

8

Appl. No. 09/845,561
Amdt. dated July 20, 2004
Reply to Office Action of April 20, 2004

the identified characteristics. For example, a set of tags may be developed to specify prosodic

characteristics similar to those of the speech of a particular speaker. These tags may be applied

to text at suitable locations within the text and may define prosodic characteristics of speech to

be generated by processing the text. The set of tags defines prosodic characteristics in sufficient

detail that processing of the tags along with the text can accurately model speech having the

prosodic characteristics of the original speech from which the tags were developed.

A set of tags can be defined by training, for example, by identifying a corpus of training

text as read by a particular speaker to identify prosodic characteristics of speech of that speaker.

Tags can be defined using the identified characteristics. Once tags have been generated, they

may be entered into a body of text for which text to speech conversion is desired. Tags may be

placed using an editor, for example, or may be placed automatically according to a programmed

set of rules. Once a body of text has been developed with a set of tags, the tags are processed to

generate speech having characteristics defined by the tags. The speech may be played using a

voice synthesizer.

The Art Rejections

All of the art rejections hinge on the application of Ladd, standing alone or Tanenblatt

and Baba standing in combination. As addressed in greater detail below, Ladd, Tanenblatt and

Baba do not support the Official Action's reading of them and the rejections based thereupon

should be reconsidered and withdrawn. Further, the Applicant does not acquiesce in the analysis

of Ladd, Tanenblatt and Baba made by the Official Action and respectfully traverses the Official

Action's analysis underlying its rejections.

9

Appl. No. 09/845,561
Amdt. dated July 20, 2004
Reply to Office Action of April 20, 2004

The Official Action rejected claims 1-9, 23, 27 and 33 under 35 U.S.C. 102(e) as unpatentable over Ladd. In light of the present amendments to claims 1 and 23, this ground of rejection is respectfully traversed.

Claim 1, as amended, claims analyzing one or more instances of actual phenomena to identify characteristics of the instances of the actual phenomena and creating a set of tags defining the identified characteristics of the one or more instances of the actual phenomena, each tag controlling one or more aspects of one or more modeled phenomena to be produced in response to the tags, the tags controlling the aspects of the modeled phenomena so as to create characteristics in the modeled phenomena similar to those exhibited by the one or more instances of the actual phenomena. These features are not taught by Ladd. Ladd teaches the use of tags to control various aspects of an interaction, including prosody of synthesized speech. However, Ladd does not teach analyzing one or more instances of actual phenomena to identify characteristics of the instances of the actual phenomena and creating a set of tags defining the identified characteristics of the one or more instances of the actual phenomena. Moreover, Ladd does not teach that the tags control aspects of modeled phenomena to be produced in response to the tags, with the tags controlling the aspects of the modeled phenomena so as to create characteristics in the modeled phenomena similar to those exhibited by the one or more instances of the actual phenomena. Ladd describes the use of a markup language, and essentially teaches the use of tags to mark text and commands so that the text, and the actions taken in response to commands, will have the desired characteristics. Ladd teaches a prosody element of a markup language allowing a user to control various aspects of content presented to a user. Attributes to

10

be controlled include rate, volume, pitch and range. The attributes can be set using numerical

values which are included in tags. See Ladd, col. 34, lines 34-58. Ladd does not suggest

analyzing instances of actual phenomena to generate tags that can produce modeled phenomena

similar to the actual phenomena, and does not present or suggest any mechanism for doing so.

Using actual phenomena to generate tags, and using the tags to produce phenomena having

characteristics similar to those of the actual phenomena, as is claimed by claim 1, presents a very

powerful and simple way of modeling phenomena. Tags generated through an analysis of actual

phenomena can be expected to produce a natural effect. For example, in text to speech

conversion, analysis of a training corpus of actual speech to produce tags can be expected to

generate a set of tags that can be used to model speech that will have a relatively natural sound.

In addition, generating tags through analysis of actual phenomena can be expected to be much

simpler than generating tags through direct intervention by a user. If a user chooses the

characteristics defined by tags, the user may be expected to do considerable experimentation in

order to achieve a desired effect. For example, if a set of tags were to be used in text to speech

conversion, the user might create tags specifying particular values for pitch or volume, place the

tags in a body of text, produce a text to speech conversion of the text and then play the speech.

The user would then make modifications, repeat the conversion and replay the speech, repeating

the process until the desired characteristics were produced. Much less, if any, such

experimentation will be required when actual phenomena are used to produce tags defining

characteristics of the actual phenomena, and those tags are used to produce modeled phenomena,

11

Appl. No. 09/845,561
Amdt. dated July 20, 2004
Reply to Office Action of April 20, 2004

as is accomplished by the invention as claimed by claim 1. Claim 1, as amended, therefore

defines over the cited art and should be allowed.

Claim 23, as amended, claims a prosody tag generation component to analyze a training

corpus to identify characteristics exhibited by one or more readings of text by one or more target

speakers and to generate a set of tags defining the identified characteristics, a text input interface

for receiving a text input and a speech modeler operative to process the text input to produce

speech having the prosodic characteristics specified by the tags, such that the speech produced by

the speech modeler is similar to that of the one or more target speakers. As noted above with

respect to claim 1, Ladd does not teach analyzing an instance of an actual phenomenon, such as a

training corpus, to identify characteristics of the phenomenon, such as characteristics of a reading

of text by a target speaker, in order to generate a set of tags defining the identified characteristics,

and does not teach producing a modeled phenomenon, such as speech, having characteristics

specified by the tags, with the modeled phenomenon having characteristics similar to those of the

actual phenomenon. Claim 23, as amended, therefore defines over the cited art and should be

allowed.

The Official Action rejected claim 21 under 35 U.S.C. 103(a) as unpatentable over

Tanenblatt in view of Baba. In light of the present amendment to claim 21, this ground of

rejection is respectfully traversed. Claim 21, as amended, claims receiving speech representing

reading of a training text by the target speaker to form a training corpus, the training corpus

representing actual sounds produced by the reading of the training text by the target speaker and

exhibiting prosodic characteristics of actual speech of the target speaker, analyzing the training

12

corpus to identify prosodic characteristics of the training corpus and creating a set of tags

defining the identified prosodic characteristics of the training corpus. These features are not

taught by Tanenblatt, Baba or a combination thereof.

Tanenblatt teaches a visual interface allowing a user to design speech characteristics for

synthesized speech. The user is presented with a mechanism for seeing and selecting various

options that are available. Once the user has made his or her choices, speech can be generated

that has the selected characteristics. The user may play the speech to determine if it is

satisfactory and make desired modifications. The system of Tanenblatt is very different from the

invention as claimed by claim 21, which uses an analysis of a training corpus representing actual

sounds produced by a reading of a training text by a target speaker in order to identify prosodic

characteristics of the training corpus, and creation of a set of tags defining the identified prosodic

characteristics of the training corpus. The creation of tags through analysis of an actual reading

of a training corpus can be expected to produce synthesized speech having relatively natural

sounding characteristics. In addition, such creation of tags can be expected to be simpler than

that of Tanenblatt, which requires a user to make actual selections of desired speech

characteristics.

Adding Baba to Tanenblatt does not cure Tanenblatt's deficiencies as a reference with

respect to claim 21, as amended. Baba teaches the use of tags to control speech characteristics of

the reading of a document, but does not teach that the tags are generated through the analysis of a

training corpus representing actual sounds produced by the reading of a training text by a target

speaker and exhibiting prosodic characteristics of actual speech of the target speaker. Rather,
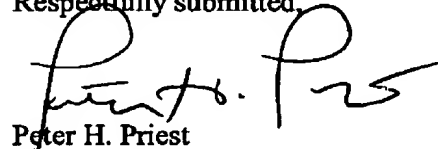
13

Tanenblatt teaches the use of tags including numerical values to set characteristics of speech, and

the selection of desired numerical values for the tags. The generation of tags as claimed by claim

21, as amended, can be expected to produce tags that will have the effect of more closely

modeling natural speech and that are simpler to generate than the tags of Baba. Claim 21, as

amended, therefore defines over the cited art and should be allowed.

Conclusion

All of the presently pending claims, as amended, appearing to define over the applied

references, withdrawal of the present rejection and prompt allowance are requested.

Respectfully submitted,

Peter H. Priest
Reg. No. 30,210
Priest & Goldstein, PLLC
5015 Southpark Drive, Suite 230
Durham, NC 27713-7736
(919) 806-1600